LABORATORIO 2

Unidad 3

Camila Paladines

Computación Científica

Profesor: Hernán Darío Vargas Cardona, PhD

Abril 9 de 2021

RESUMEN

ABSTRACT

Contenido

1. Introducción

2. Materiales y métodos

2.1. Materiales

Para el desarrollo de esta unidad se usó Python 3.7, con las librerías `numpy`, `matplotlib.pyplot`, `time`, `pandas` y `sklearn`.

2.2. Métodos

2.2.1. Ajuste de Datos

Dados m datos (t_i, y_i), se busca encontrar el vector x de parámetros que provea el mejor ajuste a una función $f(t, x)$, donde $f : \mathbb{R}^{n+1} \to \mathbb{R}$. El ajuste está dado por:

$$min\, x \sum_{i=1}^{m} (y_i - f(t_i, x))^2$$

que se conoce como la solución de mínimos cuadrados, la cual es lineal si la función f es lineal en los componentes del vector x. Esto significa que f es una combinación lineal de funciones ϕ que dependen de t:

$$f(t, x) = x_1\phi_1(t) + x_2\phi_2(t) + \cdots + x_n\phi_n(t)$$

Un ajuste lineal polinómico se puede expresar de la siguiente manera:

$$f(t, x) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$$

2.2.2. Mínimos Cuadrados Lineal

Un problema de mínimos cuadrados lineal se puede escribir en notación matricial:

$$Ax \approx b$$

Donde $A_{ij} = \phi_j(t_i)$ y $b_i = y_i$. Por ejemplo, en el ajuste polinomial cuadrático el cual tiene tres parámetros, y teniendo un conjunto de 5 datos (t_1,y_1), ...,

(t_5,y_5), entonces la matriz A sería de 5×3, y el problema tiene la forma:

$$Ax = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_1^2 \\ 1 & t_3 & t_1^3 \\ 1 & t_4 & t_1^4 \\ 1 & t_5 & t_1^5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \approx \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = b$$

2.2.3. Método de Ecuaciones Normales

Un sistema de ecuaciones normales de $n \times n$ dado por $A^\top Ax = A^\top b$ se puede usar para solucionar el problema de mínimos cuadrados lineal $Ax \approx b$.

La matriz $A^\top A$ es simétrica y positiva definida, por lo cual se puede computar su descomposición de Cholesky:

$$A^\top A = LL^\top$$

donde L es una matriz triangular inferior y $L^\top$ triangular superior. La solución x del problema de mínimos cuadrados se puede computar solucionando dos sistemas triangulares: $Ly = A^\top b$ y $L^\top x = y$.

2.2.4. Método de Householder

Mediante transformaciones Householder sucesivas, se puede introducir ceros columna por columna, debajo de la diagonal de una matriz A, para llevarla a la forma triangular. Cuando se aplica esta transformación a un vector x, se nota que:

$$Hx = \left(I - 2\frac{vv^\top}{v^\top v} \right) x = x - \left(2\frac{v^\top x}{v^\top v} v \right)$$

siendo la expresión de la derecha más barata computacionalmente y sólo requiere conocer el vector v. Este proceso sucesivo produce una factorización de la forma:

$$H_n \dots H_1 A = \begin{bmatrix} R \\ O \end{bmatrix}$$

donde R es triangular superior.

El producto de transformaciones Householder sucesivas $H_n \dots H_1$ es también una matriz ortogonal. Por lo tanto, si se toma:

$$Q^\top = H_n \dots H1,\; Q = H_1^T \dots H_n^T$$

Entonces:

$$A = Q \begin{bmatrix} R \\ O \end{bmatrix}$$

Esto quiere decir que al final se debe resolver el siguiente problema de mínimos cuadrados triangular equivalente:

$$\begin{bmatrix} R \\ O \end{bmatrix} x \approx Q^\top b$$

3. Resultados de las simulaciones

El conjunto de datos usado para la realización de las pruebas de los algoritmos fue tomado de GitHub.

3.1. Parámetros x

Los parámetros calculados ($n = 3$) por cada método se muestran en la siguiente tabla:

x	**Ecuaciones Normales**	**Householder**
x_1	295.893	203.985
x_2	-1.648	-0.423
x_3	0.007	0.004

3.2. Tiempo de cómputo

Para establecer el tiempo de cómputo de los algoritmos, se tomaron los tiempos 5 veces y se calculó el promedio, como se muestra en la siguiente tabla:

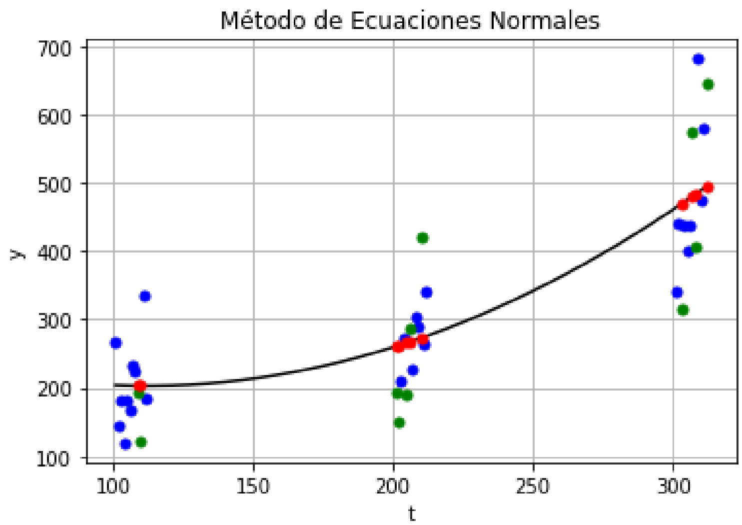
Tiempo	**Ecuaciones Normales (s)**	**Householder (s)**
1	0.00031	0.00051
2	0.00029	0.00038
3	0.00028	0.00033
4	0.00041	0.00053
5	0.0003	0.00048
Promedio	0.000318	0.000446

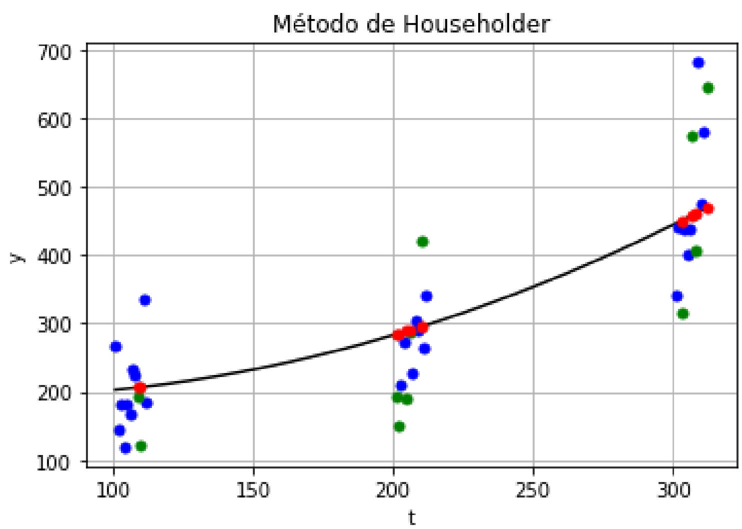
3.3. Exactitud de los algoritmos

La exactitud de los algoritmos se obtuvo mediante el Error Cuadrático Medio. En el caso del Método de Ecuaciones Normales fue de 10275.186, mientras que el del Método de Householder fue de 11385.695.

3.4. Gráfica del ajuste de mínimos cuadrados

En la siguiente gráfica se puede apreciar los resultados obtenidos por cada método. Los puntos azules corresponden a los datos de entrenamiento, los puntos verdes a los de validación, y los puntos rojos a los resultados con los parámetros calculados tomando como entradas los datos de validación.





4. Discusión y análisis

Al usar los métodos implementados en un conjunto de datos real, se pueden establecer las siguientes consideraciones:

-

5. Conclusiones

6. Referencias

- Material del curso, disponible en BlackBoard
- Bornemann, F., 2016. *Numerical linear algebra.* 1st ed. *Simson, W.*
- Mathews, J., Fink, K., Fernández Carrión, A. & Contreras Márquez, M., 2011. *Métodos Numéricos con MATLAB*. 3rd ed. Madrid: *Pearson Prentice Hall.*
- Librería Numpy
- Librería Pyplot (Matplotlib)
- Librería Time
- Librería Pandas
- Librería Scikit Learn
- Error Cuadrático Medio (ECM)